

MELHOR PREDIÇÃO LINEAR NÃO VICIADA (BLUP) DE VALORES GENÉTICOS NO MELHORAMENTO DE *Pinus*

Marcos Deon Vilela de Resende^{*}
Douglas Fabrício Prates^{**}
Andreia de Jesus^{***}
Cibele Keiko Yamada^{***}

RESUMO

O presente trabalho comparou cinco procedimentos de estimação/predição de valores genéticos no melhoramento de *Pinus*. Em situações de dados balanceados e homogeneidade de variâncias genética e ambiental, os métodos de quadrados mínimos ordinários (**OLS**), quadrados mínimos generalizados (**GLS**), melhor predição (**BP**), melhor predição linear (**BLP**) e melhor predição linear não viciada (**BLUP**) se equivalem para efeitos de ordenamento de materiais genéticos (mas não para estimação/predição de valores genéticos e ganhos genéticos). Em qualquer situação, o método **BLUP** é igual ou superior aos demais. Em geral, são melhores os seguintes métodos, em ordem decrescente: **BLUP**, **BLP**, **BP**, **GLS** e **OLS**.

PALAVRAS-CHAVE: predição de variáveis aleatórias, quadrados mínimos generalizados, quadrados mínimos ordinários, modelos lineares mistos.

BEST LINEAR UNBIASED PREDICTION (BLUP) OF BREEDING VALUES IN *Pinus* IMPROVEMENT

ABSTRACT

Five procedures of estimation/prediction of breeding values in *Pinus* improvement were compared. For balanced data with homogeneous variances, ordinary least square (**OLS**), generalized least square (**GLS**), best prediction (**BP**), best linear prediction (**BLP**) and best linear unbiased prediction (**BLUP**), best linear prediction (**BLP**) and best linear unbiased parents/ families (but not in terms of estimation/prediction of breeding values and genetic gains). In any situation, the **BLUP** method is equal or superior to the others. Generally, the best methods in descending order are: **BLUP**, **BLP**, **BP**, **GLS** and **OLS**.

* Eng. Agrônomo, Mestre, CREA n^o 50602/D, Pesquisador da Embrapa - Centro Nacional de Pesquisa de Florestas.

** Acadêmico de Estatística, Setor de Ciências Exatas, Universidade Federal do Paraná - UFPR.

*** Acadêmica de Ciência da Computação, Setor de Ciências Exatas - PUC-PR.

KEY WORDS: prediction of random variables, generalized least square, ordinary least square, mixed linear models.

1. INTRODUÇÃO

A acurada predição de valores genéticos é fundamental no melhoramento genético de espécies florestais e outras espécies perenes, de maneira geral. Entretanto, métodos diferenciados de predição são demandados em função das diferentes situações experimentais e do balanceamento associados aos dados experimentais.

Na determinação dos valores genéticos, estes podem ser considerados como efeitos fixos a serem estimados ou como variáveis aleatórias a serem preditas (SEARLE et al., 1992; WHITE & HODGE, 1989). Quando tratados como efeitos fixos, os valores genéticos são estimados, geralmente, através de médias associadas aos diferentes níveis dos efeitos fixos. Neste caso, o processo de estimação é feito pelo método dos quadrados mínimos ordinários (**OLS**) (para situações de homogeneidade de variâncias) ou pelo método dos quadrados mínimos ponderados ou generalizados (**GLS**) (para situações de heterogeneidade de variâncias). Estes processos foram aplicados por COTTERILL et al. (1983).

Tratando os valores genéticos como variáveis aleatórias, a predição pode ser efetuada pelo menos em três situações distintas: (a) iguais quantidades e precisões das informações associadas a todos os candidatos à seleção, primeiros momentos (médias) e segundos momentos (variâncias) conhecidos ou estimados com precisão; (b) diferentes quantidades e precisão das informações associadas aos candidatos à seleção, primeiros momentos e segundos momentos conhecidos ou estimados com precisão; (c) diferentes quantidades e precisões das informações associadas aos candidatos à seleção, segundos momentos conhecidos ou estimados com precisão, primeiros momentos não conhecidos ou não estimados com precisão pelo método dos quadrados mínimos ordinários.

A situação (a), cujos preditores são "melhores preditores" (**BP**), ocorre com frequência no melhoramento florestal, os seus fundamentos teóricos já são amplamente difundidos e, já existe instrumental prático (programa RESENDE et al. (1996). Os algoritmos implementados por RESENDE et al. (1996), denominados **BLP**, são similares ao procedimento denominado quadrados mínimos "regredidos" (**RLS**) relatados no melhoramento animal por HENDERSON (1948); LINDLEY & SMITH (1972); HENDERSON (1978) e HARVEY (1979).

A situação (c) exige preditores **BLUP** (melhores preditores lineares não viciados) os quais têm sido provados serem superiores aos procedimentos **RLS** (HENDERSON, 1978 e HARVEY, 1979). A vantagem do procedimento **BLUP** sobre o **BLP** é a estimação mais precisa dos efeitos fixos do modelo linear misto, a qual é realizada por **GLS** e não **OLS** como no **BLP**.

O presente trabalho tem como objetivo aplicar o procedimento **BLUP** à seleção em testes de procedência e progênie instalados em mais de um experimento e compará-lo aos métodos **OLS**, **GLS**, **BP** e **BLP**. Nesta situação, a estimação dos efeitos fixos associados às médias experimentais demanda a aplicação do método **GLS**, em função da presença de diferentes procedências e progênies em cada um dos experimentos. Isto pode conduzir a fortes influências genéticas no efeito de

experimento (o qual em essência deve ser função dos efeitos ambientais).

2. MATERIAL E MÉTODOS

2.1. Modelos lineares mistos e sua aplicação na predição de valores genéticos

Na descrição do procedimento **BLUP** faz-se necessário a descrição do modelo misto (constituído por efeitos fixos e aleatórios) associado aos dados experimentais. Conforme HENDERSON (1949; 1950; 1963; 1973; 1977; 1984; 1988; 1990) este modelo equivale a:

$$Y = X\beta + Z\mu + \Sigma, \text{ onde:}$$

Y = vetor dos dados ($N \times 1$);

X = matriz de valores conhecidos ($N \times p$), denominada matriz de incidência ou matriz de delineamento, a qual associa os elementos de β ao vetor Y ;

β = vetor desconhecido, associado aos efeitos fixos ($p \times 1$);

Z = matriz de valores conhecidos ($N \times q$), de incidência, a qual associa os elementos de μ ao vetor Y ;

μ = vetor aleatório não observável ($q \times 1$), de efeitos genéticos associados aos valores genéticos a predizer;

Σ = vetor aleatório não observável ($N \times 1$), associados aos efeitos residuais (erro experimental).

Admitindo-se μ e Σ com distribuição normal, tem-se (KENNEDY, 1981; RONNINGEN & VAN VLECK, 1985):

$$E \begin{bmatrix} Y \\ \mu \\ \Sigma \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \\ 0 \end{bmatrix}; \quad V^* \begin{bmatrix} Y \\ \mu \\ \Sigma \end{bmatrix} = \begin{bmatrix} V & ZG & R \\ G'Z & G & 0 \\ R & 0 & R \end{bmatrix}$$

onde:

V^* = Variância - Covariância.

V = $ZGZ' + R$ - matriz de variância - covariância de Y .

G = matriz de variância - covariância de μ , não singular.

R = matriz de variância - covariância de Σ , não singular.

Através deste modelo, o objetivo do melhorista é estimar β por GLS de forma:

- que seu estimador seja **BLUE** (melhor estimador linear não viciado) e, prever μ através do procedimento **BLUP**. Nesta situação, tem-se a função $\hat{\mu} = b (Y - X\hat{\beta})$, onde necessita-se determinar β para a predição dos valores genéticos especificados em μ . Conforme HENDERSON (1963) $COV(Y, \mu) = ZG$, $b = V^{-1}ZG$ e conseqüentemente:

$$\hat{\mu} = GZ'V^{-1}(Y - X\hat{\beta}) \quad (1)$$

O estimador **BLUE** (GLS) de $\hat{\beta}$ é $\beta: (X'V^{-1}X)^{-1}X'V^{-1}Y$

Henderson ET AL. (1959).

HENDERSON (1973) sugeriu um procedimento computacionalmente mais simples, o qual estima os efeitos fixos (β) e, simultaneamente, prediz os efeitos aleatórios (μ), originando as equações de modelo misto (**EMM**).

Desdobrando-se V em termos de G e R , as equações de modelo misto equivalem:

$$\begin{aligned} X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{\mu} &= X'R^{-1}Y \\ Z'R^{-1}X\hat{\beta} + (Z'R^{-1}Z + G^{-1})\hat{\mu} &= Z'R^{-1}Y \end{aligned} \quad (3)$$

Assim, a equação (1) pode ser reescrita como

$$\hat{\mu} = (Z'R^{-1}Z + G^{-1})^{-1} Z'R^{-1}(Y - X\hat{\beta}) \quad (4)$$

(HENDERSON, 1963), com $\hat{\beta}$ estimado conforme em (2).

Expressas em notação matricial e simplificadas (para situações de homogeneidade de variância genética e ambiental) em função do elemento R^{-1} (quando $R = I\sigma_c^2$, $G = A\sigma_A^2$ e $\lambda = \sigma_c^2 / \sigma_A^2$, sendo I a matriz identidade e σ_e^2 e σ_A^2 as variâncias ambiental e genética aditiva), as **EMM** equivalem a:

$$\begin{bmatrix} X'X & X'Z \\ ZX & Z'Z + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} \beta_0 \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix} \quad (5)$$

Assim sendo, conhecendo-se ou estimando-se à parte as variâncias-covariâncias (matriz V , G e R), os valores genéticos podem ser preditos utilizando-se alternativamente as equações (1) e (2); (2) e (4); (3) ou (5). DEMPFLÉ (1989) e GODDARD (1995) demonstram a equivalência dessas expressões. O modelo linear misto pode ser também multivariado e neste caso as matrizes V , G e R são estruturadas incorporando-se também covariâncias entre caracteres.

2.2. Aplicação a dados experimentais e estruturação do método BLUP

Considerou-se, neste estudo, dados referentes a um teste de procedência e progênes de *Pinus maximinoi* (material genético pertencente à rede experimental da CAMCORE - Central America and Mexico Coniferous Resources), obtidos a partir de SAMPAIO (1996). O experimento foi conduzido no delineamento de blocos casualizados com nove repetições e seis plantas por parcela, cujos detalhes são apresentados por SAMPAIO (1996). Foram avaliadas oito populações e 102 progênes de polinização aberta (meios-irmãos) em dois experimentos distintos em um mesmo local, conforme a Tabela 1.

TABELA 1. Estrutura experimental associada aos dados empregados no estudo.

Experimento 1		Experimento 2	
Procedências	Progênes	Procedências	Progênes
1	1-16	2	64-71 (+ 6 comuns)
2	17-31	3	72-82 (+ 2 comuns)
3	32-45	6	83-89
4	46-57	7	90-96
5	58-63	8	97-102

Verifica-se que a maioria do material genético é diferenciado nos dois experimentos, mas que existe uma certa fração (oito progênes em duas procedências) de materiais comuns aos dois experimentos. Assim, existe uma conexão entre os dois experimentos de forma que a aplicação do **BLUP** deverá contribuir para maior eficiência na seleção de progênes, pois as informações de um experimento serão úteis para uma mais precisa (**BLUE**) estimação dos efeitos fixos referentes à média geral e médias das procedências do outro experimento. Trabalhando-se com médias de progênes, o modelo linear misto adequado a esta situação é o seguinte:

$$\bar{Y}_{ijk..} = \mu + \ell_i + p_j + f_{k(j)} + \frac{e_{ijk.}}{b} + \frac{d_{ijk.}}{nb}, \text{ onde:}$$

μ = efeito fixo da média geral.

ℓ_i = efeito fixo do local i , $i=1,2$;

p_j = efeito fixo da procedência j , $j= 1,2 \dots 8$;

$f_{k(j)}$ = efeito aleatório da progênie K dentro da procedência j , $K=1,2 \dots F$;

$\frac{e_{ijk.}}{b}$ = efeito aleatório associado a parcelas (efeito médio em relação ao número b de blocos);

$\frac{d_{ijk.}}{nb}$ = efeito aleatório associado a indivíduo dentro de parcela (efeito médio em relação a nb , onde n é o número de plantas por parcela).

Optou-se por trabalhar a nível de médias de progênes em cada local, em função da não ocorrência de parcelas perdidas, sobrevivência alta e facilidades computacionais.

O modelo assume ausência de interação genótipos x experimento (os dois experimentos estão instalados em um mesmo local) e efeito de blocos somando zero.

Para a estimação **BLUE** de β e predição **BLUP** de μ , têm-se utilizado generalizadamente em melhoramento animal, a equação (3). Para os melhoristas de animais, a simplicidade desta equação advém do fato de que, em grande parte dos casos, as matrizes R e G são diagonais ($R = I\sigma_e^2$ e $G = I\sigma_g^2$, onde I é a matriz identidade e σ_e^2 e σ_g^2 são, respectivamente, a variância residual e a variância genética entre os materiais avaliados, para o caráter em estudo). Além de computacionalmente mais simples, este método permite a obtenção de R e G a partir do conhecimento apenas do parâmetro herdabilidade

$$\left[h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2) \right].$$

Entretanto, a aplicação em melhoramento animal (WILCOX & DELORENZO, 1983; KENNEDY & SORENSEN, 1988; HILL & MACKAY, 1989; GIANOLA & HAMMOND, 1990; HAMMOND et al., 1992), geralmente, assume homogeneidade de variâncias através dos locais de teste (ENDERSON, 1984, 1988; HILL, 1984; GARRICK & VAN VLECK, 1987) e, ausência de correlação entre todos os efeitos aleatórios, o que permite escrever $R = I\sigma_e^2$ e $G = I\sigma_g^2$. Na presente situação, estas suposições podem não ser adequadas à forma que será empregada a equação (1) para a predição dos efeitos aleatórios associados a progênies dentro de procedências.

Na composição da matriz V , os seguintes estimadores foram empregados para a obtenção das variâncias e covariâncias:

- Variâncias associadas aos efeitos aleatórios (progênie dentro de procedência):

$Var(\bar{Y}_{ijk..}) = \sigma_f^2 + \sigma_g^2 / b + \sigma_d^2 / nb$, onde σ_f^2 é a variância genética entre progênies, σ_g^2 é a variância ambiental entre as parcelas e σ_d^2 é a variância entre plantas dentro de parcela.

Os componentes de variância foram computados para cada procedência e experimento, pelo método de quadrados mínimos, de forma análoga ao método III de ENDERSON (1953).

- Covariância entre médias de diferentes famílias em um mesmo teste e em diferentes testes:

$$\begin{aligned} COV(Y_{ijk..}, Y_{i'j'k'..}) &= COV(Y_{ijk..}, Y_{i'j'k'..}) = COV(Y_{ijk..}, Y_{i'j'k'..}) \\ &= COV(Y_{ijk..}, Y_{i'j'k'..}) = 0 \end{aligned}$$

- Covariância entre médias de uma mesma família, em diferentes testes:

$$\text{COV}(Y_{ijk..}, Y_{ij'k'..}) = \sigma_f^2$$

Por sua vez a matriz G é constituída por blocos diagonais, contendo σ_f^2 ou $\frac{1}{4} \sigma_A^2$ (σ_A^2 - variância genética aditiva) equivalendo a $\text{Var}(\mu_i)$.

β foi estimado por $\hat{\beta} = (X'V^{-1}X)^{-1}(X'V^{-1}Y)^{(4)}$. Uma vez que $(X'V^{-1}X)$ é singular e, portanto, admite diversas inversas generalizadas, obteve-se uma inversa generalizada através do processo de restrição somatório zero para os níveis dos efeitos fixos de experimento e de procedência. Na realidade, o processo empregado não afeta a estimativa de $X\beta$, uma vez que $X\beta$ é função estimável, conforme conceito de estimabilidade apresentado por GRAYBILL (1976) e HENDERSON (1984).

Na presente situação, a esperança do valor genético (g) de cada parental ou progênie não é zero, uma vez que as progênies originaram-se de procedências diferentes, as quais, por definição, possuem diferentes efeitos genéticos fixos. O efeito genético de procedência é similar ao efeito de grupos, amplamente considerado em melhoramento animal (HENDERSON, 1973, 1975; THOMPSON, 1979; QUAAS & POLLAK, 1981; KENNEDY, 1991). Assim, os valores genéticos devem ser preditos como função dos efeitos fixos (β) e aleatórios (μ) advindos das equações (2) e (1). HENDERSON (1984) demonstrou que qualquer função linear de estimativas **BLUE** de efeitos fixos e predição **BLUP** de efeitos aleatórios resultam em predição **BLUP** de valores genéticos, desde que se empreguem funções estimáveis dos efeitos fixos. Assim, g pode ser predito eficientemente por:

$\hat{g} = P_p \hat{\beta}_p + C'V^{-1}(Y - X\hat{\beta})$, onde \hat{g} é o vetor de valores genéticos, P_p é a matriz de incidência associada aos efeitos genéticos fixos de procedências, $\hat{\beta}_p$ é a fração de $\hat{\beta}$, contendo apenas as estimativas para os vários níveis do efeito procedências e $C' = GZ'$

Dessa forma, os valores genéticos ao longo das diferentes procedências podem ser comparados por $\hat{g}_{ik} = \hat{\beta}_i + \hat{\mu}_{ik}$, onde $\hat{\beta}_i$ é o efeito genético da procedência i e $\hat{\mu}_{ik}$ é o efeito genético da progênie K da procedência i . Preditos desta forma, os valores genéticos e seus componentes têm as seguintes propriedades:

$$\sum_{i,k} \hat{g}_{ik} = 0; \sum_i \hat{\beta}_i = 0; \sum_k \hat{\mu}_{ik} = 0; \sum_{i,k} \hat{\mu}_{ik} = 0; \sum_k \hat{g}_{ik} = N\hat{\beta}_i$$

2.3. Comparação de vários processos de estimação/predição de valores genéticos

Foram comparados os seguintes procedimentos de estimação/predição:

- (a) Tratando os valores genéticos como efeitos fixos (constantes):
 - a.1) Método **OLS**
 - a.2) Método **GLS**
- (b) Tratando os valores genéticos como efeitos aleatórios (variáveis aleatórias):
 - b.1) Método **BP**
 - b.2) Método **BLP**
 - b.3) Método **BLUP**

Para efeito de comparação com os métodos a.1) e a.2), os métodos b.1), b.2) e b.3) não foram apresentados como desvios (em termos de \hat{g}_{ik}) em torno da média geral, mas sim somados à média geral \bar{u} , ou seja, em termos de $\hat{t} = \hat{\mu} + \hat{g}_{ik} = \hat{\mu} + \hat{\beta}_i + \hat{\mu}_{ik}$. Em termos computacionais os seguintes algoritmos foram aplicados:

$$1. \text{ OLS: } \hat{i}_1 = (X'Z)^{-1}(Z'Y)$$

$$2. \text{ GLS: } \hat{i}_2 = (Z'V^{-1}Z)^{-1}(Z'V^{-1}Y)$$

$$3. \text{ BP: } \hat{i}_3 = I_1 \sigma_g^2 Z' \left[I_2 (\hat{\sigma}_e^2 + \hat{\sigma}_g^2) \right]^{-1} (Y - X\hat{\beta}_1) + X_{up} \hat{\beta}_{1up}$$

$$4. \text{ BLP: } \hat{i}_4 = GZ'V^{-1}(Y - X\hat{\beta}_1) + X_{up} \hat{\beta}_{1up}$$

$$5. \text{ BLUP: } \hat{i}_5 = GZ'V^{-1}(Y - X\hat{\beta}_2) + X_{up} \hat{\beta}_{2up}$$

$$6. \hat{\beta}_1 = (X'X)^{-1}(X'Y)$$

$$7. \hat{\beta}_2 = (X'V^{-1}X)^{-1}(X'V^{-1}Y), \text{ onde:}$$

$\hat{\sigma}_g^2$ e $\hat{\sigma}_e^2$ - estimativas das variâncias genéticas ao nível de médias, entre as progênies avaliadas e residual respectivamente.

$X_{up}, \hat{\beta}_{1up}, \hat{\beta}_{2up}$ - matriz X e vetores $\hat{\beta}_1$ e $\hat{\beta}_2$, excluindo os efeitos fixos de locais, respectivamente.

Na computação matricial empregou-se o "software" MATLAB (MATRIX LABORATORY), desenvolvido pela Universidade de Stanford-USA (MATLAB, 1992).

3. RESULTADOS E DISCUSSÃO

Na Tabela 2 são apresentadas as estimativas referentes aos efeitos fixos da média geral, de procedências e de experimentos, empregando-se os procedimentos OLS e GLS.

TABELA 2. Estimativas* referentes aos efeitos fixos da média geral (μ), das procedências (p_i) e de experimentos (ℓ_i), obtidos pelos métodos de estimação OLS ($\hat{\beta}_1$) e GLS ($\hat{\beta}_2$), para o caráter volume de madeira em *Pinus maximinoi*.

	$\hat{\beta}_1$	$\hat{\beta}_2$
μ	0,2239	0,2240
p_1	0,2234	0,2236
p_2	0,2256	0,2296
p_3	0,2377	0,2341
p_4	0,2116	0,2118
p_5	0,1931	0,1933
p_6	0,2399	0,2398
p_7	0,2378	0,2377
p_8	0,2221	0,2221
ℓ_1	0,2460	0,2459
ℓ_2	0,2018	0,2021

* $p_i = \mu + \beta_i$, $\ell_i = \mu + e_i$, onde e_i é o efeito de experimento.

Os dois procedimentos de estimação conduziram a estimativas distintas, mas muito próximas, para todos os efeitos fixos. Isto se deve ao pequeno desbalanceamento associado ao conjunto de dados. Maiores diferenças foram obtidas para os efeitos das procedências 2 e 3, as quais possuíam progênies comuns aos dois experimentos. As melhores procedências foram **p_3 , p_6 e p_7** , obtidos pelos dois processos de estimação (Tabela 2).

Na Tabela 3 são apresentados os valores genéticos estimados/preditos através dos cinco procedimentos considerados.

TABELA3. Valores genéticos (\hat{i}) estimados/preditos através dos métodos OLS (\hat{i}_1), GLS (\hat{i}_2), BP (\hat{i}_3), BLP (\hat{i}_4) e BLUP (\hat{i}_5), para o caráter volume de madeira em 102 progênes de *Pinus maximinoi*.

Progênes	\hat{i}_1	\hat{i}_2	\hat{i}_3	\hat{i}_4	\hat{i}_5
1	0,2505	0,2505	0,2265	0,2272	0,2274
2	0,2476	0,2476	0,2247	0,2250	0,2252
3	0,2343	0,2343	0,2165	0,2149	0,2151
4	0,2235	0,2235	0,2099	0,2066	0,2068
5	0,1729	0,1729	0,1788	0,1681	0,1683
6	0,3163	0,3163	0,2668	0,2773	0,2785
7	0,2336	0,2336	0,2161	0,2143	0,2145
8	0,2383	0,2383	0,2190	0,2179	0,2181
9	0,2375	0,2375	0,2185	0,2173	0,2175
10	0,2808	0,2808	0,2450	0,2503	0,2505
11	0,2657	0,2657	0,2358	0,2387	0,2389
12	0,2199	0,2199	0,2077	0,2039	0,2041
13	0,2156	0,2156	0,2051	0,2006	0,2008
14	0,2569	0,2569	0,2304	0,2321	0,2323
15	0,2917	0,2917	0,2518	0,2586	0,2588
16	0,2432	0,2432	0,2219	0,2216	0,2218
17	0,2470	0,2470	0,2251	0,2255	0,2292
18	0,2416	0,2416	0,2218	0,2251	0,2288
19	0,2197	0,1958	0,2183	0,2197	0,2216
20	0,2714	0,2785	0,2818	0,2635	0,2654
21	0,2295	0,2295	0,2144	0,2241	0,2278
22	0,2640	0,2640	0,2355	0,2269	0,2306
23	0,2016	0,1795	0,1961	0,2110	0,2129
24	0,2709	0,2709	0,2398	0,2274	0,2311
25	0,2209	0,2209	0,2091	0,2235	0,2272
26	0,2253	0,2399	0,2252	0,2431	0,2449
27	0,2358	0,2358	0,2182	0,2246	0,2283
28	0,2340	0,2169	0,2359	0,2309	0,2327
29	0,2242	0,2013	0,2238	0,2226	0,2244
30	0,2808	0,2808	0,2459	0,2282	0,2319
31	0,1977	0,1977	0,1949	0,2216	0,2253
32	0,2975	0,2975	0,2608	0,2424	0,2393
33	0,2509	0,2432	0,2538	0,2484	0,2466
34	0,2628	0,2628	0,2395	0,2381	0,2349
35	0,2817	0,2817	0,2511	0,2404	0,2373
36	0,2650	0,2650	0,2408	0,2383	0,2352

Progenies	\hat{t}_1	\hat{t}_2	\hat{t}_3	\hat{t}_4	\hat{t}_5
37	0,2949	0,2949	0,2592	0,2421	0,2389
38	0,2394	0,2148	0,2397	0,2341	0,2323
39	0,2325	0,2325	0,2209	0,2343	0,2312
40	0,2557	0,2557	0,2351	0,2372	0,2341
41	0,2565	0,2565	0,2356	0,2373	0,2342
42	0,2646	0,2646	0,2406	0,2387	0,2352
43	0,2861	0,2861	0,2538	0,2410	0,2378
44	0,2375	0,2375	0,2240	0,2349	0,2318
45	0,2446	0,2446	0,2283	0,2358	0,2327
46	0,2369	0,2369	0,2135	0,2139	0,2141
47	0,2703	0,2703	0,2340	0,2383	0,2385
48	0,2518	0,2518	0,2227	0,2248	0,2250
49	0,2522	0,2522	0,2229	0,2251	0,2253
50	0,1975	0,1975	0,1894	0,1851	0,1853
51	0,1697	0,1697	0,1723	0,1648	0,1650
52	0,2260	0,2260	0,2069	0,2060	0,2062
53	0,2756	0,2756	0,2373	0,2422	0,2424
54	0,2104	0,2104	0,1973	0,1945	0,1947
55	0,1972	0,1972	0,1882	0,1850	0,1852
56	0,2680	0,2680	0,2326	0,2367	0,2369
57	0,2491	0,2491	0,2210	0,2228	0,2230
58	0,2787	0,2787	0,2320	0,2439	0,2441
59	0,1825	0,1825	0,1730	0,1669	0,1671
60	0,2292	0,2292	0,2017	0,2043	0,2045
61	0,2040	0,2040	0,1862	0,1841	0,1843
62	0,2151	0,2151	0,1930	0,1930	0,1932
63	0,1819	0,1819	0,1727	0,1664	0,1666
64	0,1714	0,1714	0,2059	0,2092	0,2111
65	0,2095	0,2095	0,2293	0,2286	0,2305
66	0,2124	0,2124	0,2310	0,2301	0,2320
67	0,1980	0,1980	0,2222	0,2228	0,2247
68	0,1905	0,1905	0,2176	0,2190	0,2209
69	0,1964	0,1964	0,2212	0,2219	0,2238
70	0,2423	0,2423	0,2494	0,2454	0,2473
71	0,2051	0,2051	0,2266	0,2264	0,2283
72	0,2513	0,2513	0,2596	0,2543	0,2523
73	0,1440	0,1440	0,1937	0,2045	0,2025
74	0,2305	0,2305	0,2468	0,2446	0,2426
75	0,1944	0,1944	0,2247	0,2279	0,2259
76	0,1694	0,1694	0,2093	0,2163	0,2143
77	0,1775	0,1775	0,2143	0,2200	0,2180
78	0,2357	0,2357	0,2500	0,2470	0,2450
79	0,2124	0,2124	0,2357	0,2362	0,2342

Progênes	\hat{t}_1	\hat{t}_2	\hat{t}_3	\hat{t}_4	\hat{t}_5
79	0,2124	0,2124	0,2357	0,2362	0,2342
80	0,1930	0,1930	0,2238	0,2272	0,2252
81	0,2423	0,2423	0,2541	0,2500	0,2481
82	0,2312	0,2312	0,2472	0,2449	0,2429
83	0,1881	0,1881	0,2217	0,2349	0,2348
84	0,2235	0,2235	0,2434	0,2408	0,2407
85	0,2153	0,2153	0,2383	0,2395	0,2394
86	0,2268	0,2268	0,2454	0,2414	0,2413
87	0,2512	0,2512	0,2604	0,2455	0,2454
88	0,2169	0,2169	0,2394	0,2398	0,2397
89	0,2030	0,2030	0,2308	0,2374	0,2373
90	0,2006	0,2006	0,2286	0,2317	0,2316
91	0,2485	0,2485	0,2580	0,2512	0,2511
92	0,2168	0,2168	0,2385	0,2382	0,2381
93	0,1904	0,1904	0,2223	0,2275	0,2274
94	0,1991	0,1991	0,2276	0,2310	0,2309
95	0,2273	0,2273	0,2449	0,2426	0,2425
96	0,2270	0,2270	0,2447	0,2424	0,2423
97	0,1659	0,1659	0,2011	0,2013	0,2013
98	0,2096	0,2096	0,2280	0,2279	0,2279
99	0,1930	0,1930	0,2178	0,2178	0,2178
100	0,1929	0,1929	0,2177	0,2177	0,2177
101	0,2507	0,2507	0,2532	0,2530	0,2530
102	0,1882	0,1882	0,2148	0,2149	0,2149

Dentre os estimadores/preditores de \hat{t} apresentados na Tabela 3, \hat{t}_5 é o mais preciso, pois é, conceitualmente mais completo. Assim, verifica-se que os métodos **OLS** (\hat{t}_1) e **GLS** (\hat{t}_2) superestimam os valores genéticos. Isto ocorre, porque estes procedimentos não consideram o parâmetro herdabilidade e também são inflacionados pelos efeitos fixos de experimento. Assim, não é adequado tratar os efeitos de progênes como fixos em melhoramento genético. Baseando-se em SEARLE (1987) e HUBER et al. (1992) pode-se dizer que os procedimentos **OLS**, **GLS**, **BP**, **BLP** e **BLUP** só seriam equivalentes se a quantidade de dados associados a cada acesso (progênie) tender ao infinito. Neste caso, a herdabilidade tenderia a 1,0 para todos os acessos e os procedimentos **BLP** e **BLUP** não necessitariam ser corrigidos para os efeitos fixos.

Comparando-se os métodos **OLS** e **GLS** verifica-se que existem diferenças nos valores genéticos estimados, apenas para as progênes pertencentes aos dois experimentos. Nesta situação, o método **GLS** pondera as observações de cada progênie pelo inverso da variância de cada experimento, atribuindo menores pesos às observações com maiores erros. Por outro lado, o método **OLS** efetua apenas a média aritmética simples das observações de cada progênie. Em geral, os métodos **OLS** e **GLS** só serão equivalentes se as médias de cada acesso (progênie) forem

estimadas com a mesma precisão.

Para efeito de ordenamento (não estimação de ganhos genéticos) de materiais genéticos, os métodos **OLS**, **GLS**, **BP**, **BLP** e **BLUP** serão equivalentes se as herdabilidades associadas a cada acesso forem as mesmas para todos os acessos (progênies ou candidatos à seleção). Esta situação só ocorre quando a seleção envolve apenas materiais genéticos de uma única população avaliados em um único experimento e, as médias de cada acesso são estimadas com a mesma precisão. Nas demais situações, os cinco procedimentos tendem a conduzir a ordenamentos diferentes.

Os procedimentos **BP**, **BLP** e **BLUP** regressam as predições em relação às médias (efeitos fixos) e, portanto, conduzem a menores valores genéticos que os métodos **OLS** e **GLS** (Tabela 3). Comparando-se os métodos **BP**, **BLP** e **BLUP**, verifica-se que os valores genéticos preditos assumiram diferentes valores, embora, em geral, não tenha havido grandes discrepâncias entre as predições pelos diferentes procedimentos. Isto ocorreu devido ao pequeno desbalanceamento dos dados e às variabilidades genética e ambiental, aproximadamente, homogêneas na presente situação (Tabela 3).

Pelos métodos **BP**, **BLP** e **BLUP**, os valores genéticos apresentados referem-se ao efeito de progênies e equivalem à metade do valor genético das matrizes de onde se originaram as progênies. Isto é devido ao fato de que a matriz **G** contém (1/4) da variância genética, no caso de progênies de meios-irmãos. Assim, os resultados apresentados são adequados à seleção de progênies. Para a seleção de parentais, os respectivos valores genéticos devem ser multiplicados por dois.

A adequação do método **BP** depende da validade da suposição de homogeneidade de variâncias (associadas aos diferentes candidatos à seleção) e da adequação do método **OLS** de estimação dos efeitos fixos. A adequação do método **BLP** depende apenas da adequação do método **OLS** de estimação dos efeitos fixos. Por outro lado, o método **BLUP** é completo, em geral, pois através das equações de modelo misto, simultaneamente, estima os efeitos fixos por **GLS** e prediz os efeitos aleatórios.

Assim, o método **BLUP** é melhor que o **BLP** em situações de desbalanceamento e/ou de conexão entre experimentos, situações estas em que a estimação **GLS** dos efeitos fixos é melhor que a **OLS**. Em caso de mais de um experimento, o método **BLUP** usa dados de todos os experimentos para estimar os efeitos fixos de cada um, desde que haja conexão (materiais genéticos comuns) entre os diferentes experimentos.

São computacionalmente mais simples pela ordem: **OLS**, **GLS**, **BP**, **BLP** e **BLUP**. Para efeito de ordenamento de materiais genéticos, todos os métodos podem ser equivalentes em algumas situações experimentais específicas. Para efeitos de ordenamento de materiais genéticos, predição de valores genéticos e estimação de ganhos genéticos, em geral, os melhores métodos são pela ordem: **BLUP**, **BLP**, **BP**, **GLS** e **OLS**. Assim, exceto em situação de dados balanceados (ou aproximadamente), o método **BLUP** deve ser usado.

4. CONCLUSÕES

- Em testes de procedências/progênies desbalanceados a comparação de procedências deve basear-se nos efeitos de procedências estimados pelos métodos **GLS** ou **BLUP** (estimação simultânea à predição dos efeitos aleatórios);
- Em situações de dados balanceados e homogeneidade de variância genética e ambiental os métodos **OLS**, **GLS**, **BP**, **BLP** e **BLUP** se equivalem para efeito de ordenamento de materiais genéticos visando à seleção, mas não para estimação/predição de valores genéticos e ganhos genéticos;
- Em qualquer situação o método **BLUP** é igual ou superior aos demais métodos para ordenamento de materiais genéticos, predição de valores genéticos e estimação de ganhos genéticos;
- Em geral são melhores os seguintes métodos, pela ordem: **BLUP**, **BLP**, **BP**, **GLS** e **OLS**.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- COTTERILL, P.P.; CORRELL, R.L.; BOARDMAN. Methods of estimating the average performance of families across-incomplete open-pollinated progeny tests. **Silvae Genetica**, Frankfurt, v.32, p.28-33, 1983.
- DEMPFLE, L. Estimation of breeding values. In: HILL, W.G; MACKAY, T.F.C., eds. **Evolution and animal breeding**. Wallingford: CAB International, 1989. p.181-188.
- GARRICK, D.J.; VLECK, L.D. VAN. Aspects of selection for performance in several environments with heterogeneous variances. **Journal of Animal Science**, Champaign, v.65, p.409-421, 1987.
- GIANOLA, D.; HAMMOND, K. eds. **Advances in statistical methods for genetic improvement of livestock**. Berlin: Springer-Verlag, 1990. 277p.
- GODDARD, M.E. **A summary of results in linear models, selection indices and BLUP and their proofs**. Armidale: Animal Genetics and Breeding Unit, University of New England, 1995. 18p.
- GRAYBILL, F.A. **Theory and application of the linear models**. North Scituate: Duxbury Press, 1976. 704p.
- HAMMOND, K.; GRASER, H-V.; Mc DONALD, C.A. eds. **Animal breeding: the modern approach**. University of Sydney: Post Graduate Foundation in Veterinary Science, 1992.
- HARVEY, W.R. Accuracy of simplifield regressed least-squares (SRLS) vs BLUP methods for ranking sires. **Indian Journal of Animal Genetics and Breeding**, v.1, p.7-13, 1979.

- HENDERSON, C.R. Ph.D. Tesis. **Estimation of general, specific, and maternal combining abilities in crosses among inbred lines of swine**. Ames: Iowa State University, 1948.
- HENDERSON, C.R. Estimates of changes in herd environment. **Journal of Dairy Science**, Champaign, (Abstr.), v.32, p.706, 1949.
- HENDERSON, C.R. Estimation of genetic parameters. *Ann. Math. Stat.*, (Abstra.), v.21, p.309, 1950.
- HENDERSON, C.R. Estimation of variance and covariance components. **Biometrics**, Washington, v.9, p.226-252, 1953.
- HENDERSON, C.R.; KEMPTHORNE, O.; SEARLE, S.R.; KROSIGK, C.M. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, Washington, v.15, p.192-218, 1959.
- HENDERSON, C.R. Selection index and expected genetic advance. In: HANSON, W.D.; ROBINSON, A.F. eds.. **Statistical genetics and plant breeding**, Washington: NAS-NRC, 1963, p.141-163. (NAS-NRC. Pub., 982)
- HENDERSON, C.R. Sire evaluation and genetic trends. In: **Animal Breeding and Genetics Symposium in Honor of J.Lush**. American Society of Animal Science, Champaign, Ill. 1973. p.10-41.
- HENDERSON, C.R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Washington, v.31, p.423-447, 1975.
- HENDERSON, C.R, Prediction of future records. In: POLLACK, E.; KEMPTHORNE, O.; BAILEY, I. eds. **Proceedings international conference on quantitative genetics**. Ames: Iowa State University Press, 1977, p.615-638.
- HENDERSON, C.R. Undesirable properties of regressed least squares prediction of breeding values. **Journal of Dairy Science**, Champaign, v.61, p.114-120, 1978.
- HENDERSON, C.R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462p.
- HENDERSON, C.R. Progress in statistical methods applied to quantitative genetics since 1976. In: WEIR, B.S.; EISEN, E.J.; GODDMAN, M.M.; NAMKOONG, G. eds. **Second international conference on quantitative genetics**. Sunderland: Sinauer Associates. 1988. p.85-90.
- HENDERSON, C.R. Statistical methods in animal improvement: historical overview. In: GIANOLA, D.; HAMMOND, K. eds. **Advances in statistical methods for genetic improvement of livestock**. Berlin: Springer Verlag, 1990, p.1-14.
- HILL, W.G. On selection among groups with heterogeneous variance. **Animal Production**, Edinburgh, v.39, p.473-477, 1984.
- HILL, W.G.; MACKAY, T.F.C. **Evolution and animal breeding**. Wallingford: CAB International, 1989. 313p.

- HUBER, D.A.; WHITE, T.L.; LITTELL, R.C.; HODGE, G.R. Ordinary least squares estimation of general and specific combining abilities from half-diallel mating designs. **Silvae Genetica**, Frankfurt, v.41, n.4/5, p.263-273, 1992.
- KENNEDY, B.W. Variance component estimation and prediction of breeding values. **Canadian Journal of Genetics and Cytology**, Ottawa, v.23, n.4, p.565-578, 1981.
- KENNEDY, B.W.; SORENSEN, D.A. Properties of mixed-model methods for prediction of genetic merit. In: WEIR, B.S.; EISEN, E.J.; GOODMAN, M.M.; NAMKOONG, G. eds. **Second international conference on quantitative genetics**. Sunderland: Sinauer Associates. 1988. p.91-103.
- KENNEDY, B.W. Henderson: The unfinished legacy. **Journal of Dairy Science**, Champaign, v.74, p.4067, 1991.
- LINDLEY, D.V.; SMITH, A.F. Bayes estimates for the linear model. **Journal of the Royal Statistical Society**, London, v.34, p.1-41, 1972.
- MATLAB. **High-performance numeric computation and visualization software** - reference guide. Natick: The Math Works, 1992.
- QUAAS, R.L.; POLLAK, E.J. Modified equations for sire models with groups. **Journal of Dairy Science**, Champaign, v.64, p.1868-1872, 1981.
- RESENDE, M.D.V de; HIGA, A.R.; LAVORANTI, O.J. Predição de valores genéticos no melhoramento de Eucalyptus - Melhor Predição Linear. In: 7°. CONGRESSO FLORESTAL BRASILEIRO, 7., 1993, Curitiba. **Anais**. São Paulo: SOCIEDADE BRASILEIRA SILVICULTURA. 1993. p.144-147.
- RESENDE, M.D.V de; OLIVEIRA, .E.B. de; MELINSKI, L.C.; GOULART, F.S.; OAIDA, G.R. SELEGEN -**Seleção Genética Computadorizada** - Módulo I. Best Prediction. Manual do Usuário. Curitiba. 1994, 31p.
- RESENDE, M.D.V. de; OLIVEIRA, .E.B. de; MELINSKI, L.C.; GOULART, F.S.; OAIDA, G.R. SELEGEN - **Seleção Genética Computadorizada** - Módulo II. Best Linear Prediction. Colombo: Embrapa-CNPf, 1996 (em preparação).
- RONNINGEN, K.; VAN VLECK, D. Selection index theory with practical applications. In: CHAPMAN, A.B. (ed.). **General and quantitative genetics**. Word Animal Science, A₄. Amsterdam: Elsevier. 1985. p.187-225.
- SAMPAIO, P.T.B. **Variação genética e métodos de seleção envolvendo procedências e progênes de 3 espécies de Pinus tropicais**. Curitiba: Universidade Federal do Paraná. 1996. 250p. Tese Doutorado (Em preparação).
- SEARLE, S.R. **Linear models for unbalanced data**. New York: J.Wiley, 1987.
- SEARLE, S.R.; CASELLA, G.; MCCULLOCH, C.E. **Variance components**. New York: J. Wiley, 1992. 528p.
- THOMPSON, R. Sire evaluation. **Biometrics**, Washington, v.35, p.339-353, 1979.

WHITE, T.L; HODGE, G.R. **Predicting breeding values with applications in forest tree improvement.** Kluwer, 1989. 367p.

WILCOX, C.J.; DELORENZO, M.A. Progeny testing in dairy cattle. In: INTERNATIONAL SYMPOSIUM ON ANIMAL PRODUCTION, 1983, Ribeirão Preto, p.101-113.